



Trabajo Practico Final de Modelos Grandes de Lenguaje

“Comparación de modelo LLM contra Basado en Bag of Words en análisis de sentimiento”

Carlos Montiel

CEIA – 2024

Introducción.

Estaremos verificando dos implementaciones de modelos basados en técnicas de inteligencia artificial para la inferencia de análisis de sentimiento sobre los reviews de compradores de amazon.

Por un lado constamos con un modelo basado en *“Bag of words”* llamado VADER (Valence Aware Dictionary and sEntiment Reasoner) y por el otro lado roBERTa (A Robustly Optimized BERT Pretraining Approach) basado en el modelo de Google BERT del 2018.

Desarrollo.

Analizamos primeramente el dataset el cual pesa unos 300mb y cuenta con 568.455 reviews de compradores de amazon.

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KF60	A3SGXH7AUHU8WG	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dil pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395B0RC6FGVXV	Karl	1	3	2	1309232000	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2Z27K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...

Analizamos la distribución del dataset en función de las estrellas registradas para cada review.



Luego de este pequeño análisis del dataset implementamos el `'SentimentIntensityAnalyzer'` de VADER y también roBERTa.

Utilizamos una frase de ejemplo para la comparación de las inferencias de los modelos

"This oatmeal is not good. Its mushy, soft, I don't like it. Quaker Oats is the way to go."

Con el cual registramos es siguiente resultado con VADER:

{'neg': 0.22, 'neu': 0.78, 'pos': 0.0, 'compound': -0.5448}

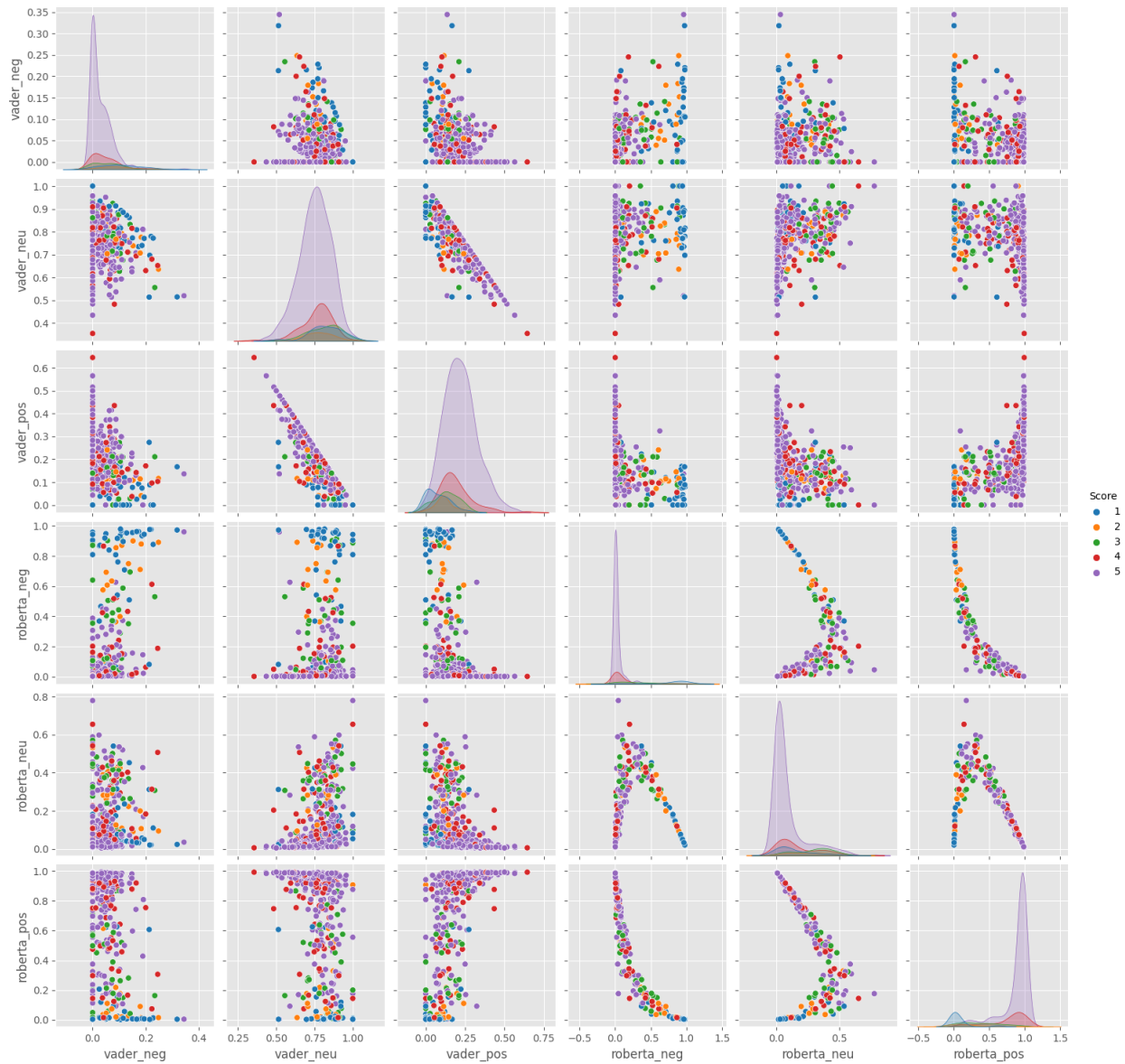
Y el siguiente resultado con roBERTa

```
{'roberta neg': 0.97635514, 'roberta neu': 0.020687465, 'roberta pos': 0.0029573692}
```

Aquí notamos diferencias en las inferencias de los diferentes modelos dando a roBERTa un scoring muy negativo a la frase de prueba y en contra partida VADER da un scoring neutro.

Resultados.

Realizamos un pairplot de los resultados obtenidos de VADER contra roBERTa, obteniendo los siguientes graficos.



También encontramos en los dos modelos casos de falsos positivos y falsos negativos

```
results_df.query('Score == 1') \
    .sort_values('roberta_pos', ascending=False)['Text'].values[0]
✓ 0.0s

'I felt energized within five minutes, but it lasted for about 45 minutes. I paid $3.99 for this drink. I could have just drunk a cup of coffee and saved my money.'
```

```
results_df.query('Score == 1') \
    .sort_values('vader_pos', ascending=False)['Text'].values[0]
✓ 0.0s

'So we cancelled the order. It was cancelled without any problem. That is a positive note...'
```

```
results_df.query('Score == 5') \
    .sort_values('roberta_neg', ascending=False)['Text'].values[0]
✓ 0.0s

'this was soooooo delicious but too bad i ate em too fast and gained 2 pds! my fault'
```

```
results_df.query('Score == 5') \
    .sort_values('vader_neg', ascending=False)['Text'].values[0]
✓ 0.0s

'this was soooooo delicious but too bad i ate em too fast and gained 2 pds! my fault'
```

Conclusiones.

Comparando los modelos propuestos llegamos a la conclusión con este dataset que los dos infieren relativamente bien los reviews de los productos salvo ciertos casos que obtenemos falsos positivos y falsos negativos con los dos modelos.

Sería interesante comparar más modelos de análisis de sentimiento como la librería Transformers la cual es relativamente fácil de utilizar y no requiere de demasiadas configuraciones.